

Confluent Cloud to Snowflake

Date published: 2021-04-06

Date modified: 2024-06-03



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

ReadyFlow: Confluent Cloud to Snowflake.....	4
Prerequisites.....	4
List of required configuration parameters for the Confluent Cloud to Snowflake ReadyFlow.....	6

ReadyFlow: Confluent Cloud to Snowflake

You can use the Confluent Cloud to Snowflake ReadyFlow to move your data from a Kafka topic hosted in a Confluent Cloud Kafka cluster to a Snowflake table.

This ReadyFlow consumes JSON, CSV or Avro data from a source Kafka topic in Confluent Cloud and parses the schema by looking up the schema name in the Confluent Schema Registry. You can filter events based on a user provided SQL query. The filtered events are then merged into CSV files, compressed into gzip format and written to the destination Snowflake DB table.

ReadyFlow details	
Source	Confluent Cloud Kafka Topic
Source Format	JSON, CSV, Avro
Destination	Snowflake DB
Destination Format	Snowflake DB

Prerequisites

Learn how to collect the information you need to deploy the Confluent Cloud to Snowflake ReadyFlow, and meet other prerequisites.

For your data ingest source

- You have created a Confluent Cloud Kafka cluster.
- You have created at least one Kafka topic.
- You have created a schema for your data in the Confluent Cloud Schema Registry.
- You have the Confluent Cloud Kafka broker endpoint.
- You have the Confluent Cloud Kafka client API Key and Secret.
- You have the Kafka Consumer Group ID.
- You have the Confluent Cloud Schema Registry endpoint.
- You have the Confluent Cloud Schema Registry API Key and Secret.

For Cloudera DataFlow

- You have enabled Cloudera DataFlow for an environment.

For information on how to enable Cloudera DataFlow for an environment, see [Enabling Cloudera DataFlow for an Environment](#).

- You have created a Machine User to use as the Cloudera Workload User.
- You have given the Cloudera Workload User the EnvironmentUser role.

1. From the Management Console, go to the environment for which Cloudera DataFlow is enabled.
2. From the Actions drop down, click Manage Access.
3. Identify the user you want to use as a Workload User.

**Note:**

The Cloudera Workload User can be a machine user or your own user name. It is best practice to create a dedicated Machine user for this.

4. Give that user EnvironmentUser role.

- You have synchronized your user to the Cloudera Public Cloud environment that you enabled for Cloudera DataFlow.

For information on how to synchronize your user to FreeIPA, see [Performing User Sync](#).

- You have granted your Cloudera user the DFCatalogAdmin and DFFlowAdmin roles to enable your user to add the ReadyFlow to the Catalog and deploy the flow definition.

1. Give a user permission to add the ReadyFlow to the Catalog.

- a. From the Management Console, click User Management.
- b. Enter the name of the user or group you wish to authorize in the Search field.
- c. Select the user or group from the list that displays.
- d. Click Roles Update Roles .
- e. From Update Roles, select DFCatalogAdmin and click Update.




Note: If the ReadyFlow is already in the Catalog, then you can give your user just the DFCatalogViewer role.

2. Give your user or group permission to deploy flow definitions.

- a. From the Management Console, click Environments to display the Environment List page.
- b. Select the environment to which you want your user or group to deploy flow definitions.
- c. Click Actions Manage Access to display the Environment Access page.
- d. Enter the name of your user or group you wish to authorize in the Search field.
- e. Select your user or group and click Update Roles.
- f. Select DFFlowAdmin from the list of roles.
- g. Click Update Roles.

3. Give your user or group access to the Project where the ReadyFlow will be deployed.

- a. Go to DataFlow Projects .
- b. Select the project where you want to manage access rights and click  More Manage Access .

4. Start typing the name of the user or group you want to add and select them from the list.

5. Select the Resource Roles you want to grant.

6. Click Update Roles.

7. Click Synchronize Users.

For your data ingest target

- You have created a Snowflake account.
- Ensure that the Snowflake database and table you are ingesting data to exist. If not, create them.
- You have set up your destination Snowflake table with the right columns and columns type.
- Your Snowflake user account has write access to the Snowflake database table.
- You have the schema of the Snowflake database. The default value is PUBLIC.
- You have created Snowflake Private and Public Keys and assigned the Public Key to your Snowflake account user.

For more information, see [Key Pair Authentication](#) in Snowflake documentation.

- You have created your Snowflake Pipe and configured the file format for CSV input.

For more information, see [CREATE FILE FORMAT](#) in Snowflake documentation.

Related Concepts

[List of required configuration parameters for the Confluent Cloud to Snowflake ReadyFlow](#)

List of required configuration parameters for the Confluent Cloud to Snowflake ReadyFlow

When deploying the Confluent Cloud to Snowflake ReadyFlow, you have to provide the following parameters. Use the information you collected in *Prerequisites*.

Table 1: Confluent Cloud to Snowflake ReadyFlow configuration parameters

Parameter Name	Description
CSV Delimiter	If your source data is CSV, specify the delimiter here.
Data Input Format	Specify the format of your input data. You can use "CSV", "JSON" or "AVRO" with this ReadyFlow. <ul style="list-style-type: none"> • CSV • JSON • AVRO
Filter Rule	Specify the filter rule expressed in SQL to filter streaming events for the destination database. Records matching the filter will be written to the destination database. The default value forwards all records.
Kafka Broker Endpoint	Specify the Kafka bootstrap server.
Kafka Client API Key	Specify the API Key to connect to the Kafka cluster.
Kafka Client API Secret	Specify the API Secret to connect to the Kafka cluster.
Kafka Consumer Group Id	The name of the consumer group used for the source topic you are consuming from.
Kafka Schema Name	Specify the schema name to be looked up in the Confluent Schema Registry for the Source Kafka Topic.
Kafka Source Topic	Specify a topic name that you want to read from.
Schema Registry Client Key	Specify the API Key to connect to the Confluent Schema Registry.
Schema Registry Client Secret	Specify the API Secret to connect to the Confluent Schema Registry.
Schema Registry Endpoint	Specify the Confluent Schema Registry API endpoint.
Snowflake Account Name	Specify the Snowflake Account Name.
Snowflake Organization	Specify the Snowflake organization name.
Snowflake Pipe	Specify the Snowflake pipe.
Snowflake Private Key File	Upload the Snowflake private key file.
Snowflake Private Key Password	Specify the password for the Snowflake private key.
Snowflake User Name	Specify your Snowflake user name.
Snowflake User Password	Specify the Snowflake user password.
Snowflake Warehouse	Specify your Snowflake warehouse. The default value is COMPUTE_WH
Table Name	Specify the Snowflake table name you want to write to.
Target Database Name	Specify the Snowflake database you want to write to.
Target Schema Name	Specify the schema of the Snowflake database. The default value is PUBLIC

Related Concepts[Prerequisites](#)**Related Information**[Deploying a ReadyFlow](#)